DOCUMENT RESUME

ED 078 006 TM 002 818

AUTHOR Barnes, Edward
TITLE IQ Testing and Minority School Children: Imperatives for Change.
INSTITUTION Connecticut Univ., Storrs. National Leadership Inst. - Teacher Education/Early Childhood.
NOTE 8p.

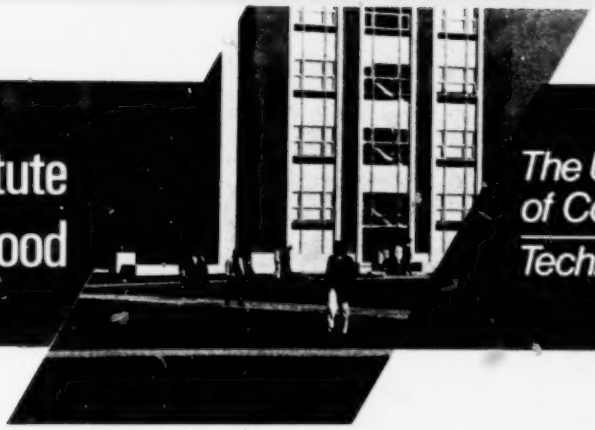
EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Ability Identification; *Culture Free Tests; Intelligence Quotient; *Intelligence Tests; *Minority Group Children; Performance Factors; Publications; Student Ability; *Test Bias; *Test Interpretation

ABSTRACT

The inadequacy and misuse of intelligence testing for minority group children are examined. IQ test items, norms, examining procedures, and language usage are discussed in terms of their bias against minority children. The implications of this bias for the classroom teacher are explored with the view that teacher mental sets are powerful mediators in learning and performance and that intelligence test scores play a major role in determining the nature of the set teachers develop. Culture specific or environment testing and criterion-referenced tests are discussed as nondiscriminatory and more valuable methods of evaluation. (KM)

National Leadership Institute
Teacher Education/Early Childhood

The University of Connecticut
Technical Paper



ED 078006

TM 002 818

IQ TESTING AND MINORITY SCHOOL CHILDREN: IMPERATIVES FOR CHANGE

Edward Barnes
University of Pittsburgh

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Edward Barnes is Associate Dean of the College of Arts and Sciences at the University of Pittsburgh. Dr. Barnes is a clinical psychologist and an expert on diagnosis and assessment procedures for atypical children and adults.

Many school systems, including New York City, Minneapolis and Philadelphia have dropped IQ tests from their testing programs. Many systems never had these tests and do not intend to add them. The National Education Association has called for a moratorium on standardized testing of minorities until tests and procedures are developed which eliminate the imprecisions of these tests for minorities, especially those children who have been mangled by deprivation and grinding poverty. Dr. Barnes' paper tells why wise school people are turning away from IQ tests for minority children and offers solid alternatives to these tests.

Decisions based on faulty IQ testing affect thousands of minority children and deny them access to quality education each year. Ideally, in the educational setting tests should be used to maximize the growth and development of the child, but for some minority children the converse is true. Instead of being tools which facilitate growth they are tools which thwart and destroy that process in these children. If you doubt this assertion ponder for a moment the following cases:

> When Juan Gonzales entered the first grade in Chicago, he was given an intelligence test and classified as mentally retarded on the basis of his score. Subsequently, after attending a special class for handicapped children for nine years, he was retested. Shortly thereafter embarrassed apologies were made to the mother by the school social worker who stated that Juan was never retarded.
>
> At about this time, in Riverside, California, Sylvia Arias was placed in a regular class after spending five years in a program for the "educable mentally retarded." A representative of the school told the father that Sylvia had been capable of doing standard school work all along. Sylvia was placed in the program for the educable mentally retarded on the basis of her IQ test score.
>
> A continent away, in Manhattan, New York, Paul Jefferson a black youngster was placed in a regular class after spending three years in a program for the mentally retarded. An embarrassed school social worker confessed to Mrs. Jefferson. "We made a mistake. This youngster was never retarded."

The rage, frustration, and helplessness felt by Mrs. Jefferson, Mrs. Gonzales, and Mr. Arias is readily understandable. But the fact is that Juan, Sylvia, and Paul were lucky ones. For them at least the stigmatizing label "mentally retarded" was eventually removed. For thousands of mis-classified and stigmatized youngsters throughout the nation vindication is not even on the horizon. This fact assumes significance when we consider the extent to which minority children are represented in educable mentally retarded (EMR) classes. Dunn (1968) stated that over 50 percent of those enrolled in classes for the retarded in this country are ethnic minority children: blacks, Chicanos, Puerto Rican Americans, and American Indians. Mercer (1971) found, in Riverside, California, that three times more Chicanos and two-and-one-half times more blacks than would be expected from their percentage in the population tested at the borderline defective or below range (a score of 79 or less) on one of the best intelligence tests in the country. Garrison and Hammill (1971) reassessing children (mostly black) placed in mentally retarded classes in the five-county-greater-Philadelphia area, found evidence which suggested that as many as two-thirds of the placements were questionable.

These cases may be considered by some to be in the extreme in that the children were placed in EMR classes. However, there are those whose scores are not low enough

FILMED FROM BEST AVAILABLE COPY

to get them into classes for the educable mentally retarded, but they are not at the "average" level (IQ 100). These students do not receive their share of attention, assistance, and reinforcement from teachers. They are not expected to learn, and as they progress through school they are subjected to subtle and not so subtle indignities, they are counseled out of high level aspirations, denied access to a college preparatory curriculum, and in various and sundry ways are forced out of the educational systems psychologically and physically.

Clearly this is an immoral, untenable and intolerable state-of-affairs as evidenced by the strong drive for reform. The black, brown, and red communities are strongly demanding a radical reform in the testing apparatus of the schools. They are demanding that tests become part of the solution rather than continue as part of the problem in the education of racial and cultural minority children. As educators and teachers, we too can demand no less. But how can reform be accomplished? A first step might be an analysis of errors in IQ testing and in the process confront squarely some of the consequences of applying these tests to minority children.

ERRORS IN IQ TEST ITEMS

This brings us to a critical group of questions: To what extent do IQ tests measure what they purport to measure? Do they measure with equal validity for all groups? If they do not, what factors influence test scores? A variety of considerations influence the answers to these questions. What the tests measure depends, among other things, on the concept of intelligence on which it is based. The notion of what intelligence is determines the nature of the behaviors built into the test. Cronback (1970), recognized as one of the foremost experts in the testing field speaking on this issue, states: "We must accept Liverant's . . . conclusion that to decide what is and what is not intelligent behavior involves a value judgment and that a person's variations in efficiency from task to task must be explained by examining his expectations and the rewards available." (p. 248)

This statement explicitly recognized the role played by values in IQ test construction. This fact has direct relevance for the oft heard charge that IQ tests are possessed of a "white middle-class bias." In these tests the behaviors tapped, item content, and style of problem solving intersect with the white middle-class experience to the virtual exclusion of the experience of many racial and cultural minorities. Jane Mercer has found an almost perfect correlation, for example, between similarities in minority family life styles to the white middle-class and scores on IQ tests. Blacks in her California sample with all five of the criterial lifestyle factors of this group match national standards on IQ tests. Those with only one factor average 82 on the tests. One upper-middle class black school led the city of Los Angeles in IQ testing in 1969.

As an illustration let us turn for a moment to the two most prestigious IQ tests in the country: The Stanford-Binet (S-B) and Wechsler Intelligence Scale for Children (WISC). A most important task in these tests asks the child to define words of increasing difficulty. Difficulty is defined in terms of frequency of use (rarity) of a given word. Now rarity is relative and depends on the language community one uses as a referent. For example, "parterre" is a rare word for the American child, but so is "singletree." However, the test developer selected "parterre" and not "singletree" as a test item. The test developers decided that rarity would be defined by reference to white middle-class experience. The child reared in a white middle-class situation is more likely to learn the meaning of "parterre" than of "singletree." If contemporary black psychologists had undertaken to construct the first intelligence test, no doubt, a different choice would have been made.

Another type of IQ test item presents a line drawing of an object with an important element missing, and requires the child to identify the missing feature. But the pictures selected are more common to the experience of the white middle-class child. One portrays a hand without fingernail polish on all fingers. Fingers *with* nail polish are not a common sight in the poor black community. Another shows a thermometer without mercury in the bulb. Thermometers are rare in the environment of the poor black child. The test does not include items based on experiences from the child's environment; for example, doors with double locks, windows with broken panes, yards without grass, etc.

A third set of test questions presents the child with some everyday problems and asks him what he would do in the situations. For example, one question asks, "What would you do if you were sent to buy a loaf of bread and the grocer said he didn't have any more?" (The only answer on which maximal credit is given is "I would go to another store"). This question rests on several assumptions, namely that there is more than one grocery store in the immediate vicinity and that it is a safe walking distance. It does not consider that out of concern for the child's safety the parents may have made it a standing rule that the child go straight to and from the store indicated, or that to go to another store might involve crossing into the territory of a gang. Nor does it consider that in some poor communities children are not sent to the store with money because of the prevalence of extortion practices. Or that credit is extended the family only by this store. Thus, it is not surprising that inner-city minority and rural children are less likely to offer the response which earns full credit. In the writer's experience, the typical response of young minority inner-city dwellers in a large mid-western city to this question is "go home," so certainly an intelligent and adaptive answer for which no credit is given.

A fourth category of items asks the child to solve some arithmetic problems. If the child has had bad teachers and bad schools and has not learned the necessary arithmetic operations - adding, subtracting, multiplying, dividing - he will be unable to solve them. If, as some people are contending, intelligence is mostly inherited, then to measure it, in part, by whether one has learned to add or not is contradictory.

In this regard, one subtest which yields minimal differences between class, ethnic, and racial groups in the United States asks the child to remember and repeat a list of four or five numbers read at a rate of one a second. It is quite

clear that specific past learning and exposure have less opportunity to operate in this instance.

Another class of IQ questions, designated analogies, asks the child to reason about concepts. The same criticisms raised with reference to the vocabulary test are relevant here. The concepts the child is expected to reason about are of differential familiarity to the various groups. Again, they are selected with reference to middle-class white experience. For example, one question asks how a "piano" and "violin" are alike, not how a "tortilla" and "frijole" are similar, or how "collards" and "sweetmilk" or "singletree" and "middle buster" are alike. Given the values operating in the selection of test items is there any wonder that the scores of whites are higher than those of blacks and other low minorities? Test developers from a socio-cultural context differing from that of the white middle-class could construct a test which would favor children from their groups.

ERRORS IN IQ TEST NORMS

Norms may be thought of as providing frames of reference for interpreting test scores - a standard so to speak. The group whose performance on the test serves as the standard is called the "normative" or "standardization" group. The performance of this group serves as the standard against which the performance of subsequent individuals taking the test is evaluated, and thereby, given meaning. Obviously then, the nature of this group is important. If the performance of a given individual is to be evaluated against a given norm, then that person should be similar to members of the normative group with respect to things which can influence performance on that test. Or to put it differently, tests should be applied only to groups and sub-classes which were included in the standardization group. For example, given the differences between the life conditions and lifeways of black youngsters and middle-class white youngsters, it is not appropriate to utilize normative data generated by the latter to evaluate the test performance of the former.

It is of interest to examine the uses of the Stanford-Binet (S-B) and the Wechsler Intelligence Scale for Children (WISC) within this framework. These are two key tests used to provide IQ scores for placement into special classes for the retarded and into other "such average" ability classes. Remember that minority youngsters are greatly disproportionately represented in "slow tracks" and educable mentally retarded classes. The S-B (1937 revision) was standardized on a sample of 3184 native born white California children of somewhat above average socio-economic status. The 1960 revision of the S-B was standardized on a sample of 4498 white children, nearly 50 percent from California suburbs. The remainder came from Minnesota, Iowa, New York, Massachusetts, and New Jersey. The south eastern, south western, and south central regions of the country were not represented in the sample. No ethnic minority children were in the standardization group. The WISC was standardized to 2200 white children from various parts of the United States. Again, no ethnic minority children were included in the standardization group. Given this fact, is it any wonder that the kinds of misclassification mentioned earlier with respect to Sylvia Arias, Juan Gonzales, and Paul Jefferson occur with such shocking frequency? Strictly speaking, the use of the S-B and WISC should be restricted to those represented in the normative group. This not only eliminates blacks, Chicanos, Puerto Ricans and Original Americans but even white children in the south eastern, south western, and south central parts of the United States. But the most important thing here is that the psychometrist who has the least semblance of training is aware of the axiom that tests should be applied only to groups which were included in the standardization population. But we see that repeatedly this axiom is violated in the use of tests with minority children.

ERRORS IN IQ EXAMINING PROCEDURES

Another source of error in IQ test scores of minority children stems from the effects of the race of the examiner on IQ scores of the examinee. In general, the evidence suggests that white examiners have subtle deleterious effects on the scores of black children. Pasamanick and Knobloch (1955) for example, found that black two year olds were inhibited in verbal expressiveness by "white" examiners. This observed verbal inhibition may be a factor in the common observation that black youngsters score higher on tests of verbal comprehension as compared to tests of verbal expression. As long ago as 1936, both white and black youngsters were observed to score higher on an IQ test when tested by members of their respective groups (Canady, 1936). Klugman (1944) found that black subjects performed better on an IQ test administered by a white examiner when they were given money incentives than when given verbal praise. White youngsters performed similarly under both conditions. Forrester and Klaus (1964) found that black kindergarteners achieved higher scores on an IQ test when examined by a black examiner than when examined by a white examiner. Other investigators have found differential responses on the part of the black adults to white and black public opinion pollsters in North Carolina (Price and Searles, 1961) and Boston Pettigrew (1964). Black pollsters elicited responses suggesting greater knowledge of current events and of the meaning of words.

Given the current mood of ethnic minorities, the negative effect produced by the white examiner on IQ tests may be heightened. For example, what is likely to be the effect of the white examiner on the black youngster when that youngster is constantly exposed to the message that white researchers are gaining fame and fortune through exploitative research of the black community, and further that psychological tests are tools of oppression of the black community when used by whites? Katz and his co-workers (Katz, et al., Katz 1964) suggest that when the administrator of an intelligence test is white, or when comparison with white peers is anticipated, black subjects perform more poorly and express concern and anxiety over their performance. An investigation by the author and two white colleagues supported the finding of Canady (1936). Repeated testing on the WISC of a sample of 13 whites and 12 black pre-teen males revealed a significant drop in average IQ score for black youngsters when tested by white examiners, a phenomenon which occurred two times within

a period of eighteen months. In several instances, items passed earlier with the black examiner were failed with white examiners. Obviously, the failure to provide the correct answer did not reflect inability to do so. Katz (1967) attempts to explain the poorer performance in the presence of the white examiner in this fashion. He hypothesizes that the anticipation of failure elicits feelings of being victimized and of covert hostility toward the tester. Since overt expression of hostility toward white authority traditionally has been fraught with danger, the impulse is suppressed and elicits emotional responses disruptive of the individual's test performance.

The negative effect of the examiner can be analyzed from another angle. White examiners generally do not come from the linguistic communities or socio-cultural backgrounds of the children tested. Kagan, the renowned Harvard child psychologist (1972), examining the test protocols of a larger number of black children from a large northeastern city, found that the children often misunderstood the examiner's pronunciation. When asked to define "fur" some responded, "that's what happens when you light a match." Obviously, the children giving this response had misinterpreted this word to be "fire" and received no test credit. Similarly, when requested to define "hat," some children answered, "when you get burned," indicating they perceived the word as "hot," and again received no credit. These are a few examples of the many ways in which the white examiner negatively impacts the black child's test performance, and renders dubious the meaning of test scores achieved.

ERRORS IN IQ TESTING RELATED TO LANGUAGE USAGE

The effect of language on test performance is clear when the child comes from a home in which a language other than English is spoken, or in which English is not spoken consistently. Intelligence tests lean heavily on verbal items and require verbal aptitude in English, thereby ignoring learning abilities in other languages. This observation is also relevant for those children who come from a language community different from the standard English community. Joan Baratz (1969) found that poor black youngsters performed better on a linguistic task involving non-standard English sentences as compared to a linguistic task involving standard English sentences. The reverse was found for lower-middle income white youngsters. Most important the black youngster performed in a superior fashion to the whites on the non-standard English sentences. Another investigator (Estelle Cherry Peisach, 1965) in an attempt to evaluate the extent to which information is successfully communicated from teacher to students of various social and cultural backgrounds and the degree of effective communication between children from different socio-cultural backgrounds, had the children restore words deleted from teachers' speech and the speech of children of diverse social backgrounds. Among other things she found that black and lower-class children did better on speech samples of children from backgrounds similar to their own. Terrell (1972) in a study conducted in the Pittsburgh area found that young lower-class black males with a low frequency of contact with middle-class individuals performed better on a task requiring the restoration of deleted words from a passage when the passage was generated by a youngster from his linguistic community. The foregoing are only a few of the many factors influencing the minority and the white child's test performance differentially. In summing up this section I quote from Guilford (1967), another recognized expert in the field of testing. He states:

> . . . That there are differences in means of test scores among racial groups, no one can deny. The meanings of these differences are not easy to determine. It can be stated as a general principle from all that we have considered with respect to conditions and their effects upon test scores, that difference among means reflect differences in needs and opportunities for the development of various kinds of abilities within the culture in which the individuals have their existence. (p. 408)

IMPLICATIONS FOR THE CLASSROOM TEACHER

Now what is the significance of all of this for the classroom teacher? Its primary significance rests on what test scores lead teachers to *believe about, feel toward, and expect of the child.* When certain things are "known" or "believed" about a child, other things, true or not, are implied. This is true for knowledge of or beliefs about IQ scores. This knowledge or belief on the part of teachers leads to the establishment of mental sets about the child to whom the IQ score is attributed. To illustrate this point three studies will be cited in detail. In the first the investigator (Cahen, 1966) was interested in investigating the import of false information regarding students' aptitudes (IQ) on teachers' scoring of students' tests. Each of 256 teachers in training was asked to score a new test of "learning readiness." Each was told that children who score higher on reading tests and on IQ tests also score higher on this new test. On the front of each test booklet the students' IQ and reading level were indicated. Sometimes these bogus scores were high, and sometimes they were low. When the teachers scored the test of the allegedly brighter children they gave them much greater benefit of the doubt than when they scored the tests of the allegedly duller children. Thus, it appears that when one "knows" a child is bright his behavior is evaluated as reflecting higher intellectual quality than is identical behavior manifested by a child "known" to be dull.

A series of investigations by Rosenthal and his coworkers approach the relationship between IQ score, teacher mental set, and evaluation of student behavior from a different angle. The investigation of interest to us involved the entire student body of an elementary school in South San Francisco (Rosenthal & Jacobson, 1966). The "Harvard Test of Inflected Acquisition" was administered to the children. This test was purported to predict academic or intellectual "blooming." The reason given for administering the test was to do a final check on its validity. In reality this test was a standardized relatively non-verbal test of intelligence, Flannagan's Test of General Ability.

In this school each of the six grades was divided into three tracks: above average, average, and below average levels of scholastic achievement. Each track was assigned to

a separate classroom. In each of the 18 classrooms, about 20 percent of the children were designated as academic "spurters." The names of these children were given to their new teacher at the start of the school year. They were told that during the academic year ahead they would show unusual intellectual gains, as suggested by scores on the test. Actually the names of that 20 percent of the children assigned to the "spurt" condition had been selected by means of a table of random numbers. Thus, the difference between those earmarked for intellectual growth, and the undesignated group of children existed only in the minds of the teachers. Four months and eight months after the teachers had been given the names of the "special" students, the test was readministered. The "special" group showed significant gains over the undesignated group in IQ score (total score and reasoning sub-scale score). This differential was particularly pronounced at early levels, grades 1 and 2. For example, at grade 1 the differential was in excess of 15 IQ points. At grade 2 it was 10 IQ points.

When teachers were asked to describe their student's classroom behavior, the "special" children were described as having a better chance of becoming successful in the future, as more interesting, curious, happy, or more appealing, and as having less need for social approval. The fascinating thing here is that these positive perceptions cannot be said to be linked to IQ score gains. A gain in IQ score when it was not expected was associated with negative teacher evaluations in the foregoing areas; that is, youngsters in the undesignated group showing substantial gains in IQ test score were rated by their teachers as less well adjusted, given to less intellectual vitality, etc. It is further interesting to note that these hazards of unpredicted intellectual growth were associated mainly with children in the low-ability tracks. This tendency toward unfavorable evaluation was observed even for the "special" lower track students. These students tend to receive less favorable evaluation than their control group peers in average and above average tracks, despite the fact that they gained as much IQ relative to the control group as did the experimental students in the highest track. In these instances apparently conflicting sets were operating, the one established by the "spurt" message and another established by the fact of low track placement. The critical factor here is that teacher expectation was associated with actual test score changes in students. Apparently, not only can scores produce mental sets in teachers, but mental sets can also produce change in actual test performance of children, even in children who are in "low ability" tracks.

The last of the three studies (Jacobson, 1966) to be cited illustrates the hypothesis that children from minority ethnic groups suffer from negative teacher attitudes. *Remember that IQ test scores produce teacher mental sets (attitudes, expectancies, beliefs, etc.).*

Two groups of teachers were asked to rank a set of unknown children's photographs on their "American" or "Mexican" appearance. ("American" was not defined.) The teachers agreed highly on their rankings. Then these same groups of teachers were asked to rank in the same manner photographs of Mexican children who were unknown to one group but were students in the school of the other group of teachers. Here there was little agreement. The teachers at the school attended by the Mexican children perceived those with higher IQs as looking more American. The significant relation of IQ and appearance was present only where the IQ scores were available. Apparently, teachers agreed in their perception of "Mexican looking" until they were made aware of the child's test score, then their perception changed.

The highest achieving (in reading), Mexican children in grades one and two were seen by both teacher groups as looking significantly more Mexican. This correlation reversed itself in grades three and four, and still more so in grades five and six; that is, the highest achievers in the upper grades looked more American to both groups of teachers. The study presented the possibility that if a Mexican child looked more American (that is, Anglo-Saxon) to a teacher, academic expectations for him might be like expectations for middle-class children as compared to those for the Mexican child who looked more Mexican, or lower-class, with resultant differences in performance.

Thus, we can assert with great confidence that teacher mental sets are powerful mediators, positive or negative, in learning and performance of children, and that intelligence test scores play a major and critical role in determining the nature of the set teachers develop. The lower scores of minority children are associated with general expectations of low cognitive performance on the part of teachers. Given these consequences of IQ tests for minority children, the implications are clear. They must be drastically changed in terms of conceptual base, makeup, administration, interpretation, and use, or they must be eliminated from the testing programs of the schools. A few attempts have been made to address the deficiencies of these tests when applied to blacks. The focus has been on test makeup, administration, and interpretation. One approach focusing on administration and interpretation involves the use of correction formulae to adjust obtained scores upward. Canady's (1936) study provided the basis for this approach. The score differential of black children when examined by white and black examiners respectively, led some to advocate adding a constant to the child's score when tested by a white examiner. Other approaches advocate using the scores on items least susceptible to a cultural bias and past learning as the best estimate of the minority child's intellectual ability. Jastec, the author of the Wide Range Achievement Test, suggests that the highest subtest score on IQ tests having subscales be taken as the index of the child's ability. Even though the use of correction formulae can reduce the number and degree of misclassifications, this approach does not get at the core of the problem. It would focus on effects of inadequate tests but leave untouched the causes and would leave the tests, and score, unchanged. Clearly this is not a satisfactory solution if solution it is at all.

THE PROMISE OF CULTURE SPECIFIC OR ENVIRONMENT TESTING

The efforts to develop "culture free," and later, "culture fair" tests were based on an apprehension of the core of the problems involved in assessing minority individuals. The concepts "culture free" and "culture fair" implicitly recognize the role of culture and learning in test behavior,

and express the ideas of rendering the test free on any culture or representing in some proportion the cultures of the groups with whom the test is to be used. Unfortunately, the effects desired of these instruments have not materialized. Several factors are responsible for the lack of impact these instruments have had on the testing apparatus of the schools. Conceptual confusion surrounds the concept "culture free," confusion having immediate implications for development of a test based on it. The problems of developing a single test which represents the several distinctive cultural groups in the country present colossal obstacles. The social context in which both culture free and culture fair tests emerged was not conducive to effecting a change in the testing conventions of the schools. The focus on the familiar, prestigious, and popular IQ tests extant precluded their acceptance.

Perhaps a potentially more fruitful approach lies in the development of "culture specific" tests. If this suggestion seems far-out, then ponder this. The model for culture specific tests already exists and when appropriately used, displays considerable effectiveness. Consider for example, the Stanford-Binet and the WISC. These are examples of "culture specific" tests. The culture in this instance is what is frequently referred to as "white middle-class." In fact, this is what the charge "white middle-class bias" refers to. But some of you will say, "but if we have different tests for different groups we will not be able to compare them in terms of intellectual ability." My rejoinder is, "but why compare them?" In what ways does an awareness of group differences in measured IQ lead to modifying the educational arrangements so that all children are effectively taught. What does the fact of a difference between groups have to say about why such differences exist? More importantly, how has the knowledge of such differences been utilized to date? The most cursory observations indicate that they have been used as a basis of pernicious labeling, a process which, as we have seen, typically leads to misclassifications and to teacher mental sets which are inimical to the learning of minority children. I must admit I become just slightly suspicious when I hear members of the majority group express an undue concern about the need to compare blacks and whites, or for that matter whites and any oppressed minority groups. Of course, you probably would grant some leeway for a slight suspicion, given the fact that over the past centuries any difference, real or fantasied, between blacks and whites has been used by whites to legitimize racist positions and practices.

If comparison there must be, then intra-group comparison where blacks, for example, can be meaningfully compared and contrasted with each other promises to be the most fruitful. Such comparisons have distinct impiications for identifying those factors which differentiate and help to condition the lives of black people within a class structured caste system. Some renowned experts in the field would argue that interracial comparison on IQ measures is desirable because this helps to identify the consequences of social deprivation and alienation. I must confess I find this position puzzling. Do we need to compare whites and blacks on IQ tests to be able to know that white racism acts as a destructive force in the lives of blacks? Parenthetically, we note once again a focus on examining the victims of racism rather than its progenitors. In any case, if group comparisons are to be made, is the IQ test score the proper medium? Of course not. Why not investigate data directly relevant to the life conditions of blacks if the goal is to assess the effects of racism? Lastly, how do we plan to use the knowledge of this difference? Presumably we would use it as a means of pointing the way to instructional approaches best adapted to the needs of minority children, to meet the child where he is, and to shape the school experience so as to maximize his development. But all readings indicate that knowledge of group differences is not used in this fashion.

Those of you who do not capitulate easily, might say, "that's all well and good but the tests predict academic performance for minority as well as for white children. In fact, they are doing what they should do as tests." I reply, "even if one grants your assertion as valid, remember that school performance is more than a matter of IQ score. It is also a function of other factors, including teacher expectancies, attitudes toward and beliefs about the student, and remember further that IQ test scores, as we have seen, structure these psychological states in teachers. The expectancies, etc., created with respect to the minority child virtually guarantee that the test will predict academic performance. You can see I am sure that the dice are loaded against the minority child, especially if he is poor."

To round out your arguments you say, "well, culture specific tests for minorities would not predict school performance as schools are currently structured, so why have them? After all, the content of such tests would bear little similarity to the content of the school curriculum, and we all know this reduces criterial validity. Furthermore, the S-B measures what is required in the school." To the first part of this statement I respond, "test building is empirical and not an endeavor proceeding on an a priori basis. So let us put it to the test." To the second part I say, "it is not a rare happening in the history of testing that the test behavior required departs markedly from criterial behavior." Thirdly, a strong case can be made for the fact that the S-B and the school curriculum share a common bias, and for the position that the curricula should reflect the cultural pluralism characterizing the society, etc. But this argument misses the heart of the matter.

The point is that "culture specific" tests could be used to determine the child's ability to function symbolically or to think in terms of his own culture and environment. After all, this is what the S-B does for the white child. *If a child can learn in one environment he can learn in another.* If a child from the Mississippi Delta has learned the relationship between "Red Bone" and "Blue Tick" or between "Sweet Milk" and "Poke Salad," or whether to run from or cook a "Tedder," that child demonstrates the same capability for conceptual thinking as the middle-class white child who has learned the relationship between "piano" and "violin." *If he can learn these relationships in his own culture, he can also master those aspects of the elementary school curriculum requiring this dimension of ability.*

Culture specific tests could be instrumental *in leading teachers to see that the content of a test is merely a*

medium for tapping mental functions; how the child thinks, perceives, and interacts with his world. The specific content of the test, when appropriate to the culture and environment of the child, eliminates those cultural, experiential, and environmental factors as determinants of performance.

Another axiom made clear to teachers would be that intelligence test performance is influenced by experience, and thus, can be taught. Development of culture specific tests for the distinctive cultural groups on the American landscape could play a major role in redirecting the nation's schools so as to make quality education available to all of the children of the society an attainable ideal.

This brings us to the second of our stated alternatives for coming to grips with the problems presented by the use of IQ tests in the school, namely the elimination of the IQ test from school testing programs. Needed reforms cannot occur so long as they occupy their central position in the testing programs of the school. Henry Dyer, Vice President of Educational Testing Services, is of the opinion that so long as IQ tests are included in the testing apparatus of the schools, needed change will be precluded. What are the arguments in favor of their elimination? What will be gained?

It is quite clear that IQ tests do not serve the needs of minority children nor of schools with heterogeneous populations. They sometimes are destructive to the interests of children in all-white schools. Yet some testing programs are still organized around IQ tests. We have seen how IQ tests are used, how they create, influence, and support beliefs and attitudes destructive in consequence on the part of those in charge of the education of minority children. We have also seen how attempts to utilize the IQ score as the point of departure for developing and organizing intervention and instructional programs have proved disastrous. Yet, the effects of major intervention programs such as Head Start are evaluated by a standard intelligence test or one very similar to it. Witness the Westinghouse evaluation of Head Start a few years ago.[2] This reflects explicitly or implicitly that it is the child's IQ that we wish to change. Without change in this domain the intervention effort is deemed not worthwhile. This focus does not concern itself with the processes of learning—how the child perceives, thinks, and interacts with his world, his pattern of strengths and weaknesses, and the relation these bear to his experience in the intervention effort. Now, obviously this is the kind of information needed if it is to be relevant to the educational endeavor. This kind of information is requisite to structuring the child's school experience to maximize his development in those areas which are or should be the focus of the educational process—discriminant analysis, convergent and divergent thinking, symbol manipulation, conservation, language development, and other skills. Tests capable of generating this kind of data would not only provide a profile for the child but could also provide specific measures of progress toward specified objectives in his instructional program. But IQ tests do not lend themselves to this kind of usage. An IQ score indicates where the child stands relative to a group (norm) with respect to his performance on some set of tasks. In this sense the IQ test is *norm referenced. It indicates little or nothing about the degree of proficiency shown by the tested behaviors in terms of what the individual can do.* The score indicates that the student is more or less proficient than another student but does not tell how proficient either is in terms of the tasks tested. *The process might be termed "child sorting." It has little educational value.*

As indicated above, other concomitants of IQ tests are inimical to the educational development of the child. Tests are sources of anxiety and threat, feelings which give rise to cheating, cramming, studying for the test on the part of the student, and teaching to the test on the part of teachers. They decree that some children must be losers so that others can be winners, and we know who the losers are going to be most of the time. In view of the foregoing it is hardly possible to escape the conclusions that the trend toward elimination of normative tests from testing programs of the schools should be accelerated.

THE PROMISE OF CRITERION-REFERENCED TESTS

Criterion-referenced tests promise to provide a real breakthrough in the struggle to ameliorate the ills of normative tests. The teaching objective is the attainment of given levels of performance in specified knowledge areas. The criterion-referenced test can address itself directly to this evaluation task because it is directed to measuring what has been taught in a particular unit by a particular teacher in a particular time span. It provides information as to the degree of competence attained by the student which is independent of reference to the performance of others. Criterion-referenced tests, in addition to providing the kind of information needed to maximize child growth and development, eliminate the need to have losers; a fact having implications for reducing motivations leading to anxiety, cheating, cramming, and other deleterious effects growing out of norm-referenced testing. Criterion-referenced tests focus on growth and behaviorally defined goals. Success is estimated in terms of the child's progress toward these goals.

The systematic denial of equal access to quality education to minority groups is an established fact. The Report of the National Advisory Commission on Civil Disobedience (1968) states that "for the community-at-large, the schools have discharged their responsibility well, but for the many minorities and particularly for the children of the racial ghetto, the schools have failed to provide the educational experience which could help overcome the effects of discrimination and deprivation" (p. 424). *The testing systems of the nation's schools play no small role in this situation.* Thus, a major change in the current picture necessitates a drastic revision in the practices and operation of these systems. Among other things, this means serious questioning of the appropriateness of the psychometric model underlying educational evaluation. Test development has been dominated by the particular requirements of predictive correlational aptitude test theory. But as we have seen, educational evaluation requires additional considerations.

Change will not come easy. Vested interests from a variety of quarters, including the testing industry and those who have traditionally advised it, and racists who would

hold on to their white supremacy illusion at all costs, will strongly resist change. But it is coming.

As educators and teachers, it is our job to see that the schools serve well all of the children of all of the people of the society. If chaos is to be averted, a commitment to the principle of pluralism in education must be effected. The ideas put forth in this paper represent some first steps in this direction. These and other needed changes can, must, and will be brought about. Failure on our part to militate for change would represent an abdication of our responsibility as professionals and as human beings, and a joining of hands with the arbiters of social chaos.

Footnotes

1. In this paper the term IQ tests is to be used synonymously with general ability tests. Not all the tests referred to in this paper are labeled IQ tests. They carry a variety of names. Some are individual tests of ability, others are group tests of ability. They all have one feature in common; they attempt to assess the child's intellectual ability. Throughout the text of this paper the Stanford-Binet and Wechsler Intelligence Scale for Children are the only tests referred to explicitly. This is because they represent the best general ability tests in the country and as such most often serve as anchor points for others and have been most widely used in the schools. They serve as the pivotal tests in the testing apparatus of the schools.

2. The Westinghouse evaluation of Head Start is a prime example of this practice. Since there was no apparent test evidence that this compensatory program produced lasting effects (increase) on the participating childrens' IQs, Head Start was deemed a failure. The implication is that since IQ is genetically determined for the most part, attempts to compensate for environmental conditions in learning are fruitless. This study was used by those in seats of power in the federal government to reduce the amount of money expended on early educational programs.

References

Baratz, Joan C. "A bi-dialectal task for determining language proficiency in economically disadvantaged Negro children." *Child Development,* 1969, *40,* 890-901.

Cahen, L. S. "Experimental manipulation of bias in teachers' scoring of subjective tests." Paper read at American Psychological Association, New York, September, 1966.

Canady, H. G. "The effects of rapport on the IQ: A new approach to the problem of racial psychology." *J. Negro Education,* 1936, *5,* 209-219.

Cronbach, L. J. *Essentials of Psychological Testing.* New York: Harper and Row, 1970.

Dunn, L. M. "Special education for the mildly retarded—Is much of it justifiable?" *Exceptional Children,* 1968, *35,* 5-22.

Forrester, B. J. & Klaus, R. A. "The effect of race of examiner on intelligence test scores of Negro kindergarten children." Peabody Papers in Human Development, 1964, *2,* 1-7.

Guilford, J. P. *The Nature of Human Intelligence.* New York: McGraw-Hill, Inc., 1967.

Jacobson, Lenore. "Explorations of variations in educational achievement among Mexican children, grades one to six." Unpublished doctoral dissertation, University of California, Berkeley, 1966.

Kagan, J. "The Magical Aura of the IQ." *Saturday Review,* Dec. 4, 1971, 92-93.

Katz, I. "Review of evidence relating to effects of desegregation on the intellectual performance of Negroes." *American Psychologist,* 1964, *19,* 381-399.

Katz, I., Epps, E. G., & Axelson, L. J. "The effects of anticipated comparison with whites and with other Negroes upon the Digit-Symbol performance of Negro college students." Unpublished manuscript." In T. F. Pettigrew, *A Profile of Negro Americans,* Princeton, N. J.: Jan Nostrand, 1964, Chapter 5.

Klugman, S. F. "The effect of money incentive vs. praise upon the reliability and obtained scores of the revised Stanford-Binet test." *J. Genetic Psychology,* 1964, *30,* 255-269.

Mercer, Jane R. "Institutionalized Anglocentrism: Labeling mental retardates in the public schools." *Race Change, and Urban Society,* 1971, *5,* 311-338.

Pasamanick, B. & Knobloch, H. "Early language behavior in Negro children and the testing of behavior." *J. Abnormal and Social Psychology,* 1955, *50,* 401-440.

Price, D. O. & Searles, Ruth. "Some effects of interviewer-respondent interaction on responses in a survey situation." Paper presented at the annual meeting of the American Statistical Association, New York, Dec. 30, 1961.

Pettigrew, T. F. *A profile of the Negro American.* New York: Van Nostrand Co., Inc., 1964.

Peisach, Estelle C., "Children's comprehension of teacher and peer speech," *Child Development,* 1965, *36,* 468-480.

Report of the National Advisory Commission on Civil Disorders, Washington, D. C.: U. S. Government Printing Office, 1968, p 424-425.

Terrell, F. "A study of Negro non-standard English between middle class black and white children who do not associate with lower class black children." Unpublished master thesis, University of Pittsburgh, 1972.